

INSTALACIÓN DE HADOOP

Iniciamos la instalación ejecutando

```
docker compose up -build
```

en base al archivo docker-compose.yml de la ruta

Una vez dentro del contenedor ejecutamos `hadoop version` para comprobar que esta correctamente instalado:

```
sh-4.2$ hadoop version
Hadoop 3.4.0
Source code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb7791783192ee7a5dfaee760
Compiled by root on 2024-03-04T06:35Z
Compiled on platform linux-x86_64
Compiled with protoc 3.21.12
From source with checksum f7fe694a3613358b38812ae9c31114e
This command was run using /opt/hadoop/share/hadoop/common/hadoop-common-3.4.0.jar
```

Listado de ficheros:

```
hdfs dfs -ls /
```

```
sh-4.2$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup          0 2024-10-25 15:36 /prueba
```

Creación de directorios:

```
sh-4.2$ hdfs dfs -mkdir /prueba
sh-4.2$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup          0 2024-10-25 15:36 /prueba
```

Copia de ficheros del sistema de ficheros local a HDFS:

```
sh-4.2$ hdfs dfs -copyFromLocal /prueba.txt /
sh-4.2$ hdfs dfs -ls /
Found 3 items
drwxr-xr-x - hadoop supergroup          0 2024-10-25 15:36 /prueba
-rw-r--r-- 1 hadoop supergroup          0 2024-10-25 15:40 /prueba.txt
drwxr-xr-x - hadoop supergroup          0 2024-10-25 15:36 /prueba2
```

Copia de ficheros desde HDFS al sistema de ficheros local:

```
sh-4.2$ hdfs dfs -copyToLocal /prueba2.txt /prueba
copyToLocal: /prueba/prueba2.txt._COPYING_ (Permission denied)
sh-4.2$ sudo chmod -R 777 /prueba
sh-4.2$ hdfs dfs -copyToLocal /prueba2.txt /prueba
sh-4.2$ ls /prueba
prueba2.txt
```

Impresión del contenido de un fichero:

```
sh-4.2$ hdfs dfs -cat /prueba.txt
Hola Hadoop
```

```
sh-4.2$ hdfs dfs -tail /prueba.txt
Hola Hadoop
```

Copiar un fichero de una ruta a otra dentro de HDFS:

```
sh-4.2$ hdfs dfs -cp /prueba.txt /prueba
sh-4.2$ hdfs dfs -ls /prueba
Found 1 items
-rw-r--r-- 1 hadoop supergroup          12 2024-10-25 16:45 /prueba/prueba.txt
```

Mover un fichero de una ruta a otra dentro de HDFS:

```
sh-4.2$ hdfs dfs -mv /prueba.txt /prueba2
sh-4.2$ hdfs dfs -ls /prueba2
Found 1 items
-rw-r--r-- 1 hadoop supergroup          12 2024-10-25 16:33 /prueba2/prueba.txt
```

Eliminar un fichero de HDFS:

```
sh-4.2$ hdfs dfs -rm /prueba/prueba.txt
Deleted /prueba/prueba.txt
```

Acceder a la ayuda para ver más comandos:

```
sh-4.2$ hdfs dfs -help
Usage: hadoop fs [generic options]
    [-appendToFile [-n] <localsrc> ... <dst>]
    [-cat [-ignoreCrc] <src> ...]
    [-checksum [-v] <src> ...]
    [-chgrp [-R] GROUP PATH...]
    [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
    [-chown [-R] [OWNER][:[GROUP]] PATH...]
    [-concat <target path> <src path> <src path> ...]
    [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
    [-copyToLocal [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
    [-count [-q] [-h] [-v] [-t <storage type>]] [-u] [-x] [-e] [-s] <path> ...]
    [-cp [-f] [-p | -p[topax]] [-d] [-t <thread count>] [-q <thread pool queue size>] <src> ... <dst>]
    [-createSnapshot <snapshotDir> [<snapshotName>]]
    [-deleteSnapshot <snapshotDir> <snapshotName>]
    [-df [-h] [<path> ...]]
    [-du [-s] [-h] [-v] [-x] <path> ...]
    [-expunge [-immediate] [-fs <path>]]
    [-find <path> ... <expression> ...]
    [-get [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
    [-getfacl [-R] <path>]
    [-getfattr [-R] {-n name | -d} [-e en] <path>]
    [-getmerge [-nl] [-skip-empty-file] <src> <localdst>]
    [-head <file>]
    [-help [cmd ...]]
    [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]
```